1968

A comparison of some parametric and non-parametric discrimination procedures in negative exponential populations

Lockett, Joseph Louis

Monterey, California. U.S. Naval Postgraduate School

A COMPARISON OF SOME PARAMETRIC AND NON-PARAMETRIC DISCRIMINATION PROCEDURES IN NEGATIVE EXPONENTIAL POPULATIONS

by

Joseph Louis Lockett

UNITED STATES
NAVAL POSTGRADUATE SCHOOL

THESIS

A COMPARISON OF SOME PARAMETRIC
AND NON-PARAMETRIC DISCRIMINATION PROCEDURES
IN NEGATIVE EXPONENTIAL POPULATIONS

by

Joseph Louis Lockett III

December 1968

A COMPARISON OF SOME PARAMETRIC AND NON-PARAMETRIC DISCRIMINATION PROCEDURES IN NEGATIVE EXPONENTIAL POPULATIONS

by

Joseph Louis Lockett III
Lieutenant, United States Navy
B.A., Rice University, 1961

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
December 1968

ABSTRACT

A brief discussion of the literature concerned with the two-population discrimination problem is presented and several procedures based on the likelihood ratio for discrimination between negative exponentially distributed populations are proposed. The small sample and asymptotic performance of these procedures is compared with that of non-parametric procedures and the classical linear discriminant function. Some guidelines for the use of the procedures discussed are presented.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

ACKNOWLEDGMENTS

I wish to express my thanks to Professor J. R. Borsting for suggesting the research which led to this thesis and for his constant assistance, inspiration, and patience. I am also grateful to Mrs. Diane Chace for her assistance with computation; and to my wife and to Mrs. Roberta Shudde who turned the rough draft into finished typescript.

I. INTRODUCTION

The problem of classification arises when one or more measurements are made on an individual and one wishes to classify the individual as belonging to one of a finite number of categories on the basis of these measurements. Each category is characterized by a probability distribution of the measurements, but the proper category of the individual is not observable; it must be inferred from the measurements. Thus the problem, in abstract terms, is: given an observation of a random variable arising from one of several populations, find a rule for deciding from which population the observation came.

The classification problem is, then, one of finding an appropriate "statistical decision function." We have a number of hypotheses: each hypothesis is that the distribution of the observation is that corresponding to a given population, and one of these hypotheses must be selected, the others rejected.

In the classification problem, there are essentially three levels of information about the distributions corresponding to the various populations which may be available to the statistician.

1. the distributions may be completely known
2. the distributions may be known to belong to a given family indexed by a parameter which is unknown
3. the distributions may be completely unknown

In cases 2) and 3), information about the value of the parameter or about the unknown distribution is usually available from a sample or sequence of realizations of the random variable corresponding to each population.

In the investigations reported in this thesis, the individual to be classified belongs to one of two populations. In this situation, case 1) above is equivalent to the simple vs. simple hypothesis testing problem whose solution is given by the Neyman-Pearson Lemma. Case 2) has received relatively little attention except under the assumption that the family of distributions is multi-variate normal with the same (but unknown) co-variance matrix. The distribution of the statistics arising in this situation have been derived. In addition, Hoel and Peterson (5) have derived very general conditions under which procedures using sample estimates of the parameters are asymptotically optimal. Case 3) was first considered by Fix and Hodges in 1951.

In Section II of this thesis the non-parametric procedure proposed by Fix and Hodges (2,3) and the application of this procedure when the distribution of the random variables is negative exponential will be reviewed. A bound on the error probabilities of the Fix-Hodges procedure discovered by Cover and Hart (1) and a more general procedure proposed by Loftsgaarden and Quesenbury (6) will also be examined.

Section III will present the results of a study of a Likelihood Ratio discrimination procedure in case 2) above and a comparison of the performance of the various procedures

considered in this thesis when the random variables have the univariate negative exponential distribution. In Section IV conclusions and recommendations arising from this study will be presented.

II. REVIEW OF LITERATURE

Notation and Definitions

In considering the classification problem, the following structure will be assumed. The two categories or populations have distribution functions F and G, and without loss of generality, since the measures with cumulative distribution functions F and G are absolutely continuous with respect to that given by F + G, the density functions f and g will be supposed to exist. Random samples from the two distributions are available: $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ independent and identically distributed as F and as G respectively; they may be used to obtain information about the respective distributions. An observation z of the random variable Z is made, and the classification problem is to decide whether Z is distributed as F or as G. The abbreviation $Z \sim F$ should be read "Z is distributed as F." The probabilities of misclassification will be designated as

$$P_1 = \Pr\{\text{assign } Z \sim G | Z \sim F\}$$

$$P_2 = \Pr\{\text{assign } Z \sim G | Z \sim G\} \quad .$$

In the case that the distributions are negative exponential,

$$F(x) = 1 - e^{-\lambda x} \text{ and } G(y) = 1 - e^{-\mu y} \quad .$$

Throughout this thesis reference will be made to discrimination procedures which tend to behave similarly in the limit; that is as the number of sample observations upon which they are based grows very large. This concept may be made explicit by introducing two notions of consistency defined by Fix and Hodges (2):

Definition 1:

The sequences of decision functions $\{\Delta_n'\}$ and $\{\Delta_n''\}$ are said to be consistent in the sense of performance characteristics if, whatever be the true distributions of the random variables, for any $\varepsilon > 0$ there exists N so that if $m \geq N$ and $n \geq N$

$$|\Pr\{\Delta_m' = \delta_i\} - \Pr\{\Delta_n'' = \delta_i\}| < \varepsilon$$

for every possible decision δ_i.

Definition 2:

The sequences of decision functions $\{\Delta_n'\}$ and $\{\Delta_n''\}$ are said to be consistent in the sense of decision functions if, whatever be the true distributions of the random variables, for any $\varepsilon > 0$, there exists N so that if $m \geq n$ and $n \geq N$

$$\Pr\{\Delta_m' = \Delta_m''\} > 1 - \varepsilon .$$

It is clear that consistency in the second sense implies that in the first. All proofs of consistency by Fix and Hodges and those in this thesis provide consistency in the stronger sense. The modifying phrase will however be omitted.

Discrimination when the distributions are completely known

When the two distributions F and G are completely known, the problem of assigning an observation z to one of the two may be posed as a test of the hypothesis $Z \sim F$ against the alternative $Z \sim G$. In this case, the Neyman-Pearson Lemma gives the procedure: Assign Z as distributed according to F if

$$\frac{f(z)}{g(z)} > t \qquad \text{where } t \text{ is to be determined } 0 \leq t \leq \infty$$

Assign $Z \sim F$ with probability γ if

$$\frac{f(z)}{g(z)} = t \quad .$$

Otherwise assign $Z \sim G$. This procedure is optimal in that for any assigned probability of error "of the first kind," i.e., $\Pr\{\text{assign } Z \sim G | Z \sim F\} = P_1$, the probability of error "of the second kind," i.e., $\Pr\{\text{assign } Z \sim F | Z \sim G\} = P_2$, of this procedure is no greater than that of any other. The value of t is chosen in the classical hypothesis-testing problem so that the probability of error of the first kind is some chosen value. Since the class of Neyman-Pearson tests is equivalent to the class of Bayes tests, the above procedure (for the appropriate choice of t) is also optimal with respect to minimizing any given weighted sum of the two error probabilities.

This procedure will be designated L(t). In the case that F and G are negative-exponential distributions, the L(t) procedure is:

$$\text{Assign } Z \sim F \text{ if and only if } \frac{\lambda}{\mu} e^{(\mu-\lambda)z} \geq t \ .$$

<u>Discrimination when the distributions are completely unknown</u>

When nothing can be assumed about the form of the distribution corresponding to the two populations, the statistician has only the observations $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ from which to obtain information enabling him to classify Z appropriately. The procedures which Fix and Hodges (2) suggest involve the estimation of the densities f and g at the point of interest, and the use of these estimates in the likelihood ratio

procedure. The following theorem due to Fix and Hodges demonstrates the asymptotic optimality of this procedure.

<u>Theorem 1</u>: Let $\hat{f}$ and $\hat{g}$ denote estimates of the densities f and g respectively and let $L^*(t;\hat{f},\hat{g})$ denote the likelihood ratio discrimination procedure using $\hat{f}$ and $\hat{g}$ in place of f and g. If $\hat{f}_{m,n}(z)$ and $\hat{g}_{m,n}(z)$ are consistent estimates for f(z) and g(z) for all z except possibly for $z \in N_{f,g}$ where $P_F(N_{f,g}) = 0 = P_G(N_{f,g})$ then $L^*_{m,n}(t;\hat{f},\hat{g})$ is consistent with L(t).

The problem, then, is reduced to that of finding consistent estimates of the densities f and g. If the observation space is reduced to one dimension by a non-negative transformation ρ, such that $x_n \to x$ entails $\rho(x_n,x) \to 0$, and if, further, for each z except possibly for a null set under both the F and G distribution $\rho(X,z)$ and $\rho(Y,z)$ are random variables with continuous densities not both zero at zero, then given the observation z to be classified, the observations $X_1,\ldots,X_m$; $Y_1,\ldots,Y_n$ may be replaced by $\rho(X_1,z),\ldots, \rho(X_m,z)$; $\rho(Y_1,z),\ldots, \rho(Y_n,z)$ and the discrimination involves non-negative univariate random variables. A consistent estimate of the transformed densities is given by the following theorem of Fix and Hodges.

<u>Theorem 2</u>: Let X and Y be non-negative. Let f and g be positive and continuous at 0. Let k(m,n) be a positive, integer-valued function such that $k(m,n) \to \infty$, $\frac{1}{m} k(m,n) \to 0$ and $\frac{1}{n} k(m,n) \to 0$ as $m,n \to \infty$ with $\frac{m}{n} \to \delta \neq 0$ or ∞. Define

$U = k^{th}$ smallest value of the combined samples of X's and Y's

M = number of X's $\leq U$

N = number of Y's $\leq U$

then $\frac{M}{nU}$ is a consistent estimate for $f(0)$ and $\frac{N}{nU}$ is a consistent estimate for $g(0)$.

The $L^*(t,\hat{f},\hat{g})$ procedure thus requires: Assign $Z \sim F$ if and only if

$$\frac{\hat{f}}{\hat{g}} = \frac{M/m}{N/n} \geq t \ .$$

<u>Performance of the Non-Parametric Discriminator with finite samples</u>

Fix and Hodges (3) continued the investigation of their non-parametric discrimination procedure by examining its performance for small samples where distributions are Normal with identical covariance matrix; that is, under conditions in which the linear discriminant function is known to be an optimal procedure. The bulk of that investigation is for univariate distributions with k (the total number of the available samples used in the classification) equal one. This is the "Rule of Nearest Neighbor": classify $Z \sim F$ if and only if z's nearest neighbor is an x. Fix and Hodges obtain the misclassification probability for this procedure for a considerable range of sample sizes and for distance between population means of 1, 2 and 3 times the standard deviation. Limiting error probabilities (as $m = n \to \infty$) are obtained for $k = 1$ and $k = 3$ with distance between population means of 1 to 5 times

the standard deviation. Some results are obtained for bivariate normal distributions and an estimate of the performance of the discriminator for $k > 3$ is obtained. One very interesting result of this investigation is that, regardless of the underlying distributions, as $m = n \to \infty$ the two error probabilities of the rule of nearest neighbor are equal and no greater than one-half.

Hager (4) investigated the performance of the "rule of nearest neighbor" under the assumption that F and G were negative exponential. He contrasted this with the performance under the same conditions, of the linear discriminant function and obtained misclassification probabilities for a wide range of (equal) sample sizes and parameter values for the latter procedure when F and G were Gamma distributions of order 1 to 20. His results in the exponential case are included in Section III of this thesis.

Loftsgaarden and Quesenbury (6) proposed an alternative density estimator to that suggested by Fix and Hodges, which is consistent and applicable in a Euclidean space of any dimension. The procedure is let $j(m)$ be a sequence of integers such that

$$\lim_{m\to\infty} j(m) = \infty$$

$$\lim_{m\to\infty} \frac{j(m)}{m} = 0 \quad .$$

To estimate the density at a point z, using a sample $x_1,\ldots,x_m$, let $w_{(1)},\ldots,w_{(m)}$ be the transformed sample $|x_1-z|,\ldots,|x_m-z|$ ordered from smallest to largest. Let $A_{w(j),z}$ denote

the volume (Lebesgue measure) of the hypersphere of radius $w_{(j)}$ centered at z, then

$$\tilde{f}(z) = \frac{j-1}{n} \cdot \frac{1}{A_{w(j),z}}$$

is a consistent estimate of the density f at the point z.

If the density g at z is similarly estimated based on $y_1,\ldots,y_n$, denoting the transformed sample by $v(1),\ldots,v(\ell),\ldots,v(n)$ (where $\ell(n)$ is a sequence with the same characteristics as j above), then by Theorem 1 the procedure $L^*(t;\tilde{f},\tilde{g})$ which requires, assign $Z \sim F$ if and only if

$$\frac{\frac{j-1}{m} \cdot \frac{1}{A_{w(j),z}}}{\frac{\ell-1}{n} \cdot \frac{1}{A_{v(\ell),z}}} \geq t$$

is consistent with the procedure L(t) and hence asymptotically optimal. Note that, if t = 1 and m = n, j = ℓ, this procedure is identical with the Fix-Hodges procedure with $k = j + \ell - 1$ since a majority of the k nearest neighbors of z are x's if and only if $w(j) < v(\ell)$. In the general case, the procedures $L^*(t;\hat{f},\hat{g})$ and $L^*(t;\tilde{f},\tilde{g})$ are quite similar but not identical. The density estimate $\tilde{f}$ has applicability to problems other than that of classification, while the estimate $\hat{f}$ is not so versatile.

In their paper, Loftsgaarden and Quesenbury report a small empirical study of the density estimator $\tilde{f}$ when the true distributions are Uniform, negative exponential, and Normal. Based on this study, they recommend that the sequence j(n) take values not less than $n^{\frac{1}{2}}$.

In an article published in 1967, Cover and Hart (1) evaluated the rule of nearest neighbor in a slightly different context from that in which the previous investigations had placed it. Their work is in a Bayesian context so that there is a probability structure over the space {F,G}

$$\eta_1 = Pr\{Z \sim F\}$$

$$\eta_2 = Pr\{Z \sim G\} \quad .$$

It is assumed also that the random sample of X's and Y's arise in a way so that there is one fixed sample size with the number of X's within that sample being probabilistically determined.

If the classification loss function simply counts wrong decisions, i.e., the loss is 0 or 1 depending on whether the observation to be classified is assigned correctly or incorrectly; if R* designates the expected risk of the Bayes procedure with respect to a given prior distribution $(\eta, 1-\eta)$ where $\eta = Pr\ \{Z \sim F\}$ and if R designates the expected risk (with respect to the same prior distribution) of the rule of nearest neighbor, then the result for discrimination between two populations proved by Cover and Hart is given by the following:

<u>Theorem 3</u>: Let the space of possible values of the random variables be a separable metric space. Let f and g be such that, with probability one x is either 1) a continuity point of f and g, or 2) a point of non-zero probability measure.

Then the expected risk R of the nearest neighbor procedure has the bounds

$$R^* \leq R \leq 2R^*(1-R^*) \quad .$$

These bounds are as tight as possible.

A comparable bound is obtained for the case of discrimination among several populations.

III. A LIKELIHOOD RATIO DISCRIMINANT

As was noted in the last section, when the probability structure of the two populations to be discriminated is known completely the likelihood ratio criterion gives the solution to the classification problem: that is, classify z as distributed according to F if

$$\frac{f(z)}{g(z)} \geq t \quad \text{for some } t,\ 0 \leq t \leq \infty \ .$$

The procedure which Fix and Hodges selected with which to compare the rule of nearest neighbor was the linear discriminant function, since that procedure is known to be optimal under the assumption that the populations under consideration are Normally distributed with the same covariance matrix. Investigation of the linear discriminant reveals that it is the likelihood ratio procedure using the estimates of the population means and the common co-variance matrix as though they were known to be correct. Hager's investigation indicated that the use of the linear discriminant when the populations have the negative exponential distribution can give very poor results and that, in general, the probability of misclassification is divided very unevenly between P_1 and P_2. It is not surprising that the linear discriminant performs poorly on distributions so radically different from the Normal as the negative exponential. In fact, good performance in this case would be quite surprising.

In attempting to discover a parametric discrimination procedure with good properties, one might emulate the development which leads to the discriminant function and suggest that

the random sample of the two populations be used to estimate the parameter of the distributions. The likelihood ratio procedure could then be carried out as though the estimates were known to be correct. This procedure which will be designated $L(t;\hat{\lambda},\hat{\mu})$ would then be

$$\text{Let } \hat{\lambda} = \frac{m}{\sum_{i=1}^{m} X_i} \qquad \hat{\mu} = \frac{n}{\sum_{i=1}^{n} Y_i}$$

Assign $Z \sim F$ if

$$\frac{\hat{\lambda}}{\hat{\mu}} e^{(\hat{\mu}-\hat{\lambda})z} \geq t \quad \text{for some } t \quad 0 \geq t \geq \infty .$$

One may easily verify that this procedure is, indeed, asymptotically optimal. Since $\hat{\lambda} \xrightarrow{P} \lambda$ and $\hat{\mu} \xrightarrow{P} \mu$ as $n,m \to \infty$, this result follows from Theorem 4 below, or from a more general theorem of Hoel and Peterson (5).

<u>Theorem 4</u> (Fix and Hodges): If

a) the estimates $\{\hat{\theta}_{m,n}\}$ are consistent and

b) for every θ, $f_\theta(z)$ and $g_\theta(z)$ are continuous functions of θ for every z except perhaps for $z \in N_\theta$ where $Pr(N_\theta) = 0$ under the distribution given by f_θ and that given by g_θ, then the sequence of discrimination procedures obtained by applying the likelihood ratio principle with critical value $t > 0$ to $f_{\hat{\theta}_{m,n}}(z)$ and $g_{\theta_{m,n}}(z)$ is consistent with $L(t)$.

It is noteworthy that the foregoing procedure (and the linear discriminant function as well) makes no use of the observation z in determining the estimates of the parameters.

One might suppose that the use of z for this purpose would improve the performance of the procedure, at least for small sample sizes. Accordingly one could pose the problem as one of testing the composite hypothesis H_0: $z \sim F$ against the alternative H_1: $z \sim G$, using the maximum likelihood estimates λ and μ in both cases so that

$$\tilde{\lambda} = \frac{(m+1)}{\sum_{i=1}^{m} x_i + z} \; ; \; \tilde{\mu} = \frac{(n+1)}{\sum_{i=1}^{n} y_i + z} \quad .$$

Accept H_0 if

$$\frac{\tilde{\lambda}}{\tilde{\mu}} \; e^{(\tilde{\mu}-\tilde{\lambda})z} \geq t \quad .$$

This procedure which will be called $L(t;\tilde{\lambda},\tilde{\mu})$ is, of course, asymptotically equivalent to $L(t;\hat{\lambda},\hat{\mu})$, so that it too is consistent with $L(t)$ and hence optimal in the limit.

In the discussion up to this point, the problem of the choice of t in the two families of procedures which have been proposed has not been considered. The following lemma will clarify the problem.

Lemma 1: If t is restricted to be a constant in the procedure $L(t;\hat{\lambda},\hat{\mu})$ or $L(t;\tilde{\lambda},\tilde{\mu})$ as λ, μ range over the parameter space, then if $t \neq 1$, as $m,n \to \infty$ for any $\varepsilon > 0$ there exists δ so that if $|1 - \frac{\mu}{\lambda}| < \delta$, $P_1 > 1 - \varepsilon$ or $P_2 > 1 - \varepsilon$.

Proof: The procedure $L(t;\hat{\lambda},\hat{\mu})$ requires: assign $Z \sim F$ iff

$$\frac{\hat{\lambda}}{\hat{\mu}} \; e^{(\hat{\mu}-\hat{\lambda})z} \geq t$$

or

$$(\hat{\mu}-\hat{\lambda})z \geq \ln\frac{\hat{\mu}}{\hat{\lambda}} + \ln t \quad .$$

Let $m,n \to \infty$ so that $\hat{\lambda} \overset{P}{\to} \lambda$, $\hat{\mu} \overset{P}{\to} \mu$ and suppose (without loss of generality) that $\lambda < \mu$. Then the procedure assigns $Z \sim G$ incorrectly if and only if

$$z < \frac{\ln\frac{\mu}{\lambda}}{\mu-\lambda} + \frac{\ln t}{\mu-\lambda} \quad .$$

Now suppose $t > 1$; since

$$P_1 = \Pr\{\text{assign } Z \sim G | Z \sim F\} = \Pr\{Z < \frac{\ln\frac{\mu}{\lambda}}{\mu-\lambda} + \frac{\ln t}{\mu-\lambda}\}$$

$$= 1 - \exp\left\{\frac{\ln\frac{\mu}{\lambda}}{1-\frac{\mu}{\lambda}} + \frac{\ln t}{1-\frac{\mu}{\lambda}}\right\}$$

$$= 1 - \left(\frac{t\mu}{\lambda}\right)^{1-\frac{\mu}{\lambda}}$$

the desired inequality is achieved if

$$\left(\frac{t\mu}{\lambda}\right)^{1-\frac{\mu}{\lambda}} < \varepsilon$$

or

$$\frac{\mu}{\lambda}\left(\frac{1}{\varepsilon}\right)^{\frac{\mu}{\lambda}} < \frac{1}{t\varepsilon}$$

since, by assumption, $\frac{\mu}{\lambda} > 1$ and $t > 1$, there exists $\delta > 0$ so that if $\frac{\mu}{\lambda} - 1 < \delta$ the inequality above is satisfied and the desired conclusion follows. If $t < 1$ a similar argument shows that for appropriate values of $\frac{\mu}{\lambda}$, $P_2 > 1 - \varepsilon$. Since $L(t;\tilde{\lambda},\tilde{\mu})$ is asymptotically equivalent to $L(t,\hat{\lambda},\hat{\mu})$ the result follows for the former procedure as well.

It is noteworthy that for t = 1 as $m,n \to \infty$

$$\lim_{\frac{\mu}{\lambda}\to 1+} P_1 = \lim_{\frac{\mu}{\lambda}\to 1+} 1 - \exp\left\{\frac{\ell n \frac{\mu}{\lambda}}{1-\frac{\mu}{\lambda}}\right\}$$

$$= 1 - e^{-1}$$

and similarly

$$\lim_{\frac{\mu}{\lambda}\to 1+} P_2 = e^{-1} \quad .$$

(The subscripts on the P's are reversed as $\frac{\mu}{\lambda}\to 1^-$)

In fact, it is easily verified that, for t = 1 as $m,n \to \infty$ if $\frac{1}{2} < \frac{\mu}{\lambda} < 2$ either P_1 or P_2 is greater than one-half. This is not a desirable situation; however, it is better than the situation which obtains in the use of the linear discriminant function where, as Hager discovered, for $.3863 \approx [2(\ell n2) - 1] < \frac{\mu}{\lambda} < 1/[2(\ell n2) - 1] \approx 2.589$, $P_1 > \frac{1}{2}$ or $P_2 > \frac{1}{2}$. Recall that the rule of nearest neighbor has both error probabilities bounded above by $\frac{1}{2}$ as $m,n \to \infty$ irrespective of the population distributions.

The above results are asymptotic and imply little about the performance of the procedures for small samples. They do, however, sharpen the problem which must be faced in using the $L(t;\hat{\lambda},\hat{\mu})$ procedure. Either t is fixed at 1 (for if $t \neq 1$ the procedure may become arbitrarily bad as $m,n \to \infty$) or t is made a function of the observations. If the latter course is elected, one might be interested in preventing the possibility of misclassifying an observation with higher probability than one-half. A plausible way to pursue this goal would be to

seek a minimax procedure; i.e., one which would make P_1 equal to P_2. To do this one would, given the estimates $\hat{\lambda},\hat{\mu}$ seek $t = t(\hat{\lambda},\hat{\mu})$ so that

$$P_F\{Z: \frac{\hat{\lambda}}{\hat{\mu}} e^{(\hat{\mu}-\hat{\lambda})Z} < t\} = P_G\{Z: \frac{\hat{\lambda}}{\hat{\mu}} e^{(\hat{\mu}-\hat{\lambda})Z} > t\}$$

and use this value of t for the discrimination. The performance of this "minimax" procedure is reported in this thesis.

In the foregoing material, the ratio $\frac{\lambda}{\mu}$ has occurred frequently. It would be desirable for a discrimination procedure to depend on the parameters of the distributions only through this ratio. Indeed this is true, for both $L(t;\hat{\lambda},\hat{\mu})$ and $L(t;\tilde{\lambda},\tilde{\mu})$.

<u>Theorem 5</u>: In the procedures $L(t;\hat{\lambda},\hat{\mu})$ and $L(t;\tilde{\lambda},\tilde{\mu})$, $P_1 = \Pr\{\text{assign } Z \sim G | Z \sim F\}$ depends on λ,μ only through $c = \frac{\lambda}{\mu}$.

A lemma will be established first:

<u>Lemma 2</u>: If X has the negative exponential distribution with parameter λ, then X is distributed as $(-\ell n\ U)/\lambda$ where U has the Uniform (0,1) distribution.

Proof of lemma:

$$\Pr\{X \leq x\} = F(x) = 1 - e^{-\lambda x}$$

$$\begin{aligned} \Pr\{\frac{-\ell n\ U}{\lambda} \leq x\} &= \Pr\{\ell n\ U \geq -\lambda x\} \\ &= \Pr\{U \geq e^{-\lambda x}\} \\ &= 1 - e^{-\lambda x} \quad . \end{aligned}$$

The result follows by the Caratheodory extension theorem.

Proof of Theorem 5

Suppose $n = m = 1$. Then for the procedure $L(t;\hat{\lambda},\hat{\mu})$

$$\Pr\{\text{assign } z \sim G \mid z \sim F\} = \Pr\{\frac{Y}{X}\exp\left[(\frac{1}{Y}-\frac{1}{X})Z\right] \le t \mid Z \sim F\}$$

$$= \Pr\{\frac{\lambda\ \ell n\ V}{\mu\ \ell n\ U} \exp\left[(\frac{\mu}{\ell n\ V}-\frac{\lambda}{\ell n\ U})\frac{(-\ell n\ W)}{\lambda}\right] < t\}$$

$$= \Pr\{c\frac{\ell n\ V}{\ell n\ U} \exp\left[(\frac{1/c}{\ell n\ V}-\frac{1}{\ell n\ U})(-\ell n\ W)\right] < t\}$$

where U, V, and W are independent and identically uniformly distributed on (0,1) by lemma 2.

Similarly for $L(t;\tilde{\lambda},\tilde{\mu})$,

$$\Pr\{\text{assign } Z \sim G \mid Z \sim F\} = \Pr\{\frac{Y+Z}{X+Z} \exp\left[(\frac{2}{X+Z} - \frac{2}{Y+Z})Z\right] < t \mid Z \sim F\}$$

$$= \Pr\left\{\frac{\frac{\ell n\ V}{\mu}+\frac{\ell n\ W}{\lambda}}{\frac{\ell n\ U}{\lambda}+\frac{\ell n\ W}{\lambda}} \exp\left[\left(\frac{2}{\frac{\ell n\ U}{\lambda}+\frac{\ell n\ W}{\lambda}} - \frac{2}{\frac{\ell n\ V}{\mu}+\frac{\ell n\ W}{\lambda}}\right)\frac{(-\ell n\ W)}{\lambda}\right] < t\right\}$$

$$= \Pr\left\{\frac{c\ell n\ V+\ell n\ W}{\ell n\ U+\ell n\ W} \exp\left[\left(\frac{2}{\ell n\ U+\ell n\ W} - \frac{2}{c\ell n\ V+\ell n\ W}\right)(-\ell n\ W)\right] < t\right\}$$

The result for arbitrary m,n follows by induction.

Note that $P_2 = \Pr\{\text{assign } Z \sim F \mid Z \sim G\}$

$$= \Pr\{\frac{\hat{\lambda}}{\hat{\mu}} \exp[(\hat{\mu}-\hat{\lambda})Z] > t \mid Z \sim G\}$$

$$= \Pr\{\frac{\hat{\mu}}{\hat{\lambda}} \exp[(\hat{\lambda}-\hat{\mu})Z] < 1/t \mid Z \sim G\}$$

is equal to P_1 for the situation in which λ and μ have been interchanged and t replaced by 1/t, i.e., P_2 for $L(t;\hat{\lambda},\hat{\mu})$ equals P_1 for $L(1/t;\hat{\mu},\hat{\lambda})$. A similar statement is valid for $L(t;\tilde{\lambda},\tilde{\mu})$.

In seeking the error probabilities of the procedure $L(t;\hat{\lambda},\hat{\mu})$ and $L(t;\tilde{\lambda},\tilde{\mu})$ one must calculate

$$P_1 = P\{\frac{\hat{\lambda}}{\hat{\mu}} e^{(\hat{\mu}-\hat{\lambda})Z} < t | Z \sim F\}$$

$$= P\{\ell n \frac{1}{\hat{\mu}} - \ell n \frac{1}{\hat{\lambda}} + (\hat{\mu}-\hat{\lambda})Z < \ell n\ t | Z \sim F\}$$

where in procedure $L(t;\hat{\lambda},\hat{\mu})$, $\frac{m}{\hat{\lambda}} \sim$ Gamma (λ,m), $\frac{n}{\hat{\mu}} \sim$ Gamma (μ,n), $z \sim$ Gamma $(\lambda,1)$, and in procedure $L(t;\tilde{\lambda},\tilde{\mu})$, $\frac{m+1}{\hat{\lambda}} = U + Z$ where $U \sim$ Gamma (λ,m), $Z \sim$ Gamma $(\lambda,1)$ so that $U + Z \sim$ Gamma $(\lambda,m+1)$, $\frac{n+1}{\hat{\mu}} = V + Z$ where $V \sim$ Gamma (μ,n). In the $L(t;\hat{\lambda},\hat{\mu})$ procedure, if t is a constant, it appears that P_1 should be calculable by a straightforward triple numeral integration. In the $L(t;\tilde{\lambda},\tilde{\mu})$ procedure the boundary of the region of integration for Z involves the solution of a transcendental equation, but this too may be done numerically and P_1 calculated for fixed t. However, when t is permitted to be a function of the observations, the integral becomes intractable. For this reason, and because the investigator wished to compare the performance of the Likelihood Ratio procedures to that of the Loftsgaarden-Quesenbury procedure which is almost impossible to assess analytically, the decision was made to conduct this investigation through a Monte-Carlo study. The following procedures were investigated

1) $L(t;\tilde{\lambda},\tilde{\mu})$ $t = 1$
2) $L(t;\hat{\lambda},\hat{\mu})$ $t = 1$
3) $L(t;\tilde{\lambda},\tilde{\mu})$ "minimax"
4) $L(t;\hat{\lambda},\hat{\mu})$ "minimax"
5) Rule of nearest neighbor
6) Loftsgaarden-Quesenbury procedure $L^*(t;\tilde{f},\tilde{g})$ $t = 1$

The computer program, run on an IBM 360 computer, generated, by means of the probability integral transform, the random sample of X's and Y's, and the observation Z to be classified. The various classification procedures were performed and correct or incorrect classification of z was recorded.

The Monte Carlo procedure may be viewed as an attempt to estimate the parameter p of a Bernoulli random variable; i.e., the probability with which a randomly selected observation will be misclassified. As such, the distribution of the estimates which have been obtained may be estimated. Since p is reasonable close to one-half in all cases, and since 10,000 replications of the Monte Carlo procedure were summed, it may be assumed that the estimate

$$p = \frac{\sum_{i=1}^{10,000} B_i}{10,000} ,$$

where $B_i = 0$ with probability $(1-p)$, 1 with probability p, has approximately the Normal distribution with mean p and variance $\frac{p(1-p)}{10,000} \leq .25 \times 10^{-4}$. Hence a 95% confidence interval may be formed for the value of p in each case

$$\begin{aligned} .95 &= \Pr\{|\hat{p}-p| \leq 1.96\sigma\} \\ &\leq \Pr\{|\hat{p}-p| \leq .0098\} \quad . \end{aligned}$$

For comparison with these results, the analytically computed misclassification probabilities of the rule of nearest neighbor and linear discriminant function obtained by Hager are reproduced.

Table 1

Misclassification Error Probabilities for Procedures

Description of Procedures:

1. $L(t; \tilde{\lambda}, \tilde{\mu})$ $t = 1$

2. $L(t; \hat{\lambda}, \hat{\mu})$ $t = 1$

3. $L(t; \tilde{\lambda}, \tilde{\mu})$ "minimax"

4. $L(t; \hat{\lambda}, \hat{\mu})$ "minimax"

5. $L^*(t; \hat{f}, \hat{g})$ $t = 1$ "Rule of Nearest Neighbor"

6. $L^*(t; \tilde{f}, \tilde{g})$ $t = 1$ Loftsgaarden and Quesenbury Procedure
   $j(n) = \ell(n) = n^{\frac{1}{2}}$

7. "Rule of Nearest Neighbor" - from Hager (4)

8. Linear Discriminant Function - from Hager (4)

$C = \lambda/\mu$

$N =$ size of sample from each population upon which classification procedure is based

Table 1 (Continued)

PROCEDURE 1

N C	1.5		2.0		3.0		5.0		10.0		20.0	
	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2
1	.468	.509	.419	.510	.361	.476	.272	.432	.163	.352	.098	.290
2	.456	.507	.404	.492	.318	.438	.209	.368	.113	.279	.049	.217
5	.419	.521	.338	.482	.249	.418	.160	.336	.086	.246	.039	.170
10	.379	.528	.300	.485	.216	.414	.139	.331	.080	.237	.037	.159
20	.344	.544	.266	.489	.206	.416	.138	.330	.081	.229	.040	.144
∞	.296	.556	.250	.500	.192	.423	.134	.331	.077	.226	.043	.146

PROCEDURE 2

N C	1.5		2.0		3.0		5.0		10.0		20.0	
	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2
1	.495	.487	.465	.469	.434	.416	.371	.345	.300	.240	.261	.154
2	.467	.498	.429	.471	.356	.405	.269	.319	.200	.211	.147	.135
5	.425	.517	.347	.473	.263	.403	.185	.315	.120	.218	.076	.138
10	.381	.527	.303	.481	.224	.408	.151	.322	.099	.220	.057	.145
20	.344	.543	.267	.487	.209	.413	.144	.325	.088	.222	.049	.136
∞	.296	.556	.250	.500	.192	.423	.134	.331	.077	.226	.043	.146

Table 1 (Continued)

PROCEDURE 3

N C	1.5		2.0		3.0		5.0		10.0		20.0	
	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2
1	.518	.471	.489	.435	.474	.380	.403	.307	.296	.221	.205	.163
2	.502	.460	.482	.420	.434	.339	.350	.258	.248	.174	.155	.124
5	.494	.459	.445	.388	.374	.301	.294	.231	.197	.163	.116	.112
10	.468	.443	.427	.376	.348	.304	.265	.233	.182	.164	.111	.112
20	.457	.439	.393	.370	.333	.314	.251	.238	.172	.162	.108	.101
∞	.430	.430	.382	.382	.318	.318	.245	.245	.165	.165	.106	.106

PROCEDURE 4

N C	1.5		2.0		3.0		5.0		10.0		20.0	
	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2
1	.519	.471	.493	.440	.484	.379	.432	.301	.371	.200	.329	.127
2	.499	.466	.475	.423	.430	.342	.358	.256	.287	.161	.223	.099
5	.484	.465	.436	.393	.366	.306	.291	.233	.210	.157	.145	.103
10	.461	.447	.419	.382	.344	.308	.264	.234	.189	.160	.125	.106
20	.453	.443	.389	.373	.330	.317	.251	.238	.177	.160	.114	.098
∞	.430	.430	.382	.382	.318	.318	.245	.245	.165	.165	.106	.106

Table 1 (Continued)

PROCEDURE 5

N C	1.5		2.0		3.0		5.0		10.0		20.0	
	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2
1	.452	.522	.397	.539	.332	.518	.237	.493	.132	.437	.077	.385
2	.464	.512	.418	.502	.356	.471	.268	.411	.172	.325	.096	.272
5	.477	.497	.446	.466	.385	.418	.296	.343	.198	.257	.121	.185
10	.483	.483	.443	.464	.382	.399	.310	.318	.209	.234	.133	.158
20	.480	.481	.452	.453	.398	.398	.308	.325	.221	.222	.141	.142

PROCEDURE 6

N C	1.5		2.0		3.0		5.0		10.0		20.0	
	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2
1	.452	.522	.397	.539	.332	.518	.237	.493	.132	.437	.077	.385
2	.464	.512	.418	.502	.356	.471	.268	.411	.172	.325	.096	.272
5	.451	.510	.390	.497	.302	.454	.200	.392	.098	.311	.038	.260
10	.446	.510	.377	.477	.286	.436	.178	.360	.083	.278	.027	.220
20	.437	.503	.379	.462	.279	.420	.176	.341	.089	.247	.033	.177
50	****	****	****	****	.240	.413	.153	.329	.082	.243	.036	.168
100	****	****	****	****	.234	.423	.151	.337	.083	.237	.038	.158
∞	.296	.556	.250	.500	.192	.423	.134	.331	.077	.226	.043	.146

Table 1 (Continued)

PROCEDURE 7

N C	1.5		2.0		3.0		5.0		10.0		20.0	
	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2
1	.448	.529	.400	.533	.326	.521	.236	.487	.138	.433	.076	.391
2	.462	.510	.422	.500	.356	.467	.269	.407	.168	.328	.096	.271
5	.474	.495	.440	.472	.380	.421	.299	.345	.198	.253	.121	.185
10	.478	.488	.446	.461	.389	.405	.310	.327	.211	.232	.133	.159
20	****	****	.449	.455	.392	.398	.315	.321	.217	.225	.140	.149
∞	.481	.481	.451	.451	.395	.395	.319	.319	.222	.222	.145	.145

PROCEDURE 8

N C	1.5		2.0		3.0		5.0		10.0		20.0	
	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2
1	.448	.529	.400	.533	.326	.521	.236	.487	.138	.433	.076	.391
2	.429	.531	.364	.530	.265	.504	.157	.458	.063	.408	.021	.381
5	.397	.538	.306	.526	.190	.486	.086	.442	.020	.410	.003	.394
10	.365	.545	.264	.524	.156	.482	.066	.444	.011	.416	.001	.401
20	.331	.553	.237	.525	.144	.483	.058	.448	.007	.419	.000	.405
∞	.286	.565	.223	.528	.135	.483	.050	.451	.004	.423	.000	.408

Table 2

EXPECTED RISK FOR ALL PROCEDURES WITH PRIOR (.5, .5)

Procedure		1	2	3	4	5	6	7	8
C	N								
1.5	1	.488	.491	.495	.495	.487	.487	.488	.488
	2	.482	.483	.481	.483	.488	.488	.486	.480
	5	.470	.471	.476	.475	.487	.480	.484	.467
	10	.453	.454	.455	.454	.483	.478	.483	.455
	20	.444	.443	.448	.448	.480	.470	****	.442
	∞	.426	.426	.430	.430	****	.426	.481	.426
2.0	1	.465	.467	.462	.466	.468	.468	.467	.467
	2	.448	.450	.451	.449	.460	.460	.461	.447
	5	.410	.410	.417	.414	.456	.444	.456	.416
	10	.393	.392	.401	.400	.453	.427	.453	.394
	20	.377	.377	.381	.381	.453	.420	.452	.381
	∞	.375	.375	.382	.382	****	.375	.451	.375
3.0	1	.419	.425	.427	.431	.425	.425	.424	.424
	2	.378	.380	.386	.386	.414	.414	.411	.385
	5	.333	.333	.337	.336	.401	.378	.401	.338
	10	.315	.316	.326	.326	.391	.361	.397	.319
	20	.311	.311	.324	.323	.398	.350	.395	.313
	∞	.308	.308	.318	.318	****	.308	.395	.309

Table 2 (Continued)

EXPECTED RISK FOR ALL PROCEDURES WITH PRIOR (.5, .5)

Procedure C	N	1	2	3	4	5	6	7	8
5.0	1	.352	.358	.355	.366	.365	.365	.361	.361
	2	.289	.294	.304	.307	.340	.340	.338	.307
	5	.248	.250	.262	.262	.320	.296	.322	.264
	10	.235	.237	.249	.249	.314	.269	.319	.255
	20	.234	.235	.245	.245	.316	.258	.318	.253
	∞	.233	.233	.245	.245	****	.233	.319	.250
10.0	1	.258	.270	.259	.285	.285	.285	.286	.286
	2	.196	.206	.211	.224	.249	.249	.248	.235
	5	.166	.169	.180	.184	.227	.205	.226	.215
	10	.158	.160	.173	.175	.222	.181	.222	.213
	20	.155	.155	.167	.168	.222	.168	.221	.213
	∞	.152	.152	.165	.165	****	.152	.222	.214
20.0	1	.194	.208	.184	.228	.231	.231	.233	.233
	2	.133	.141	.140	.161	.184	.184	.184	.201
	5	.105	.107	.114	.124	.153	.149	.153	.198
	10	.098	.101	.111	.115	.145	.123	.146	.201
	20	.092	.093	.104	.106	.142	.105	.145	.202
	∞	.094	.094	.106	.106	****	.094	.145	.204

IV. SUMMARY AND CONCLUSIONS

A number of interesting facts are evident from inspection of the results of the investigations conducted in this thesis. Perhaps the most startling is that for values of c not greater than 5 and all sample sizes up to and including 20 the expected risk (with prior $(\frac{1}{2},\frac{1}{2})$) of the linear discriminant function is uniformly smaller than that for either of the non-parametric procedures (see Figure 1). The linear discriminant is equivalent to procedure $L(t;\hat{\lambda},\hat{\mu})$ with t chosen in a somewhat bizarre fashion, since it divides the positive line into two intervals which are acceptance regions for $\{Z \sim F\}$ and $\{Z \sim G\}$. Hence the linear discriminant minimizes P_2 for the P_1 which it achieves, and though the division of the total error probability is very uneven, the average is small enough to better the non-parametric procedures.

Also interesting is the fact that the expected risks of procedures $L(t;\hat{\lambda},\hat{\mu})$ and $L(t;\tilde{\lambda},\tilde{\mu})$ are almost identical even for very small sample sizes. In general P_1 is larger for $L(t;\hat{\lambda},\hat{\mu})$ than for $L(t;\tilde{\lambda},\tilde{\mu})$ but P_2 for the latter procedure is smaller so as to keep the average almost constant. The "minimax" procedure appears to achieve the desired equalization of P_1 and P_2 fairly well for moderate sample sizes ($n \geq 10$), but fails quite badly for n = 1 or 2. It appears that, for $n \geq 5$ the average risk is not increased appreciably by using the "minimax" procedure.

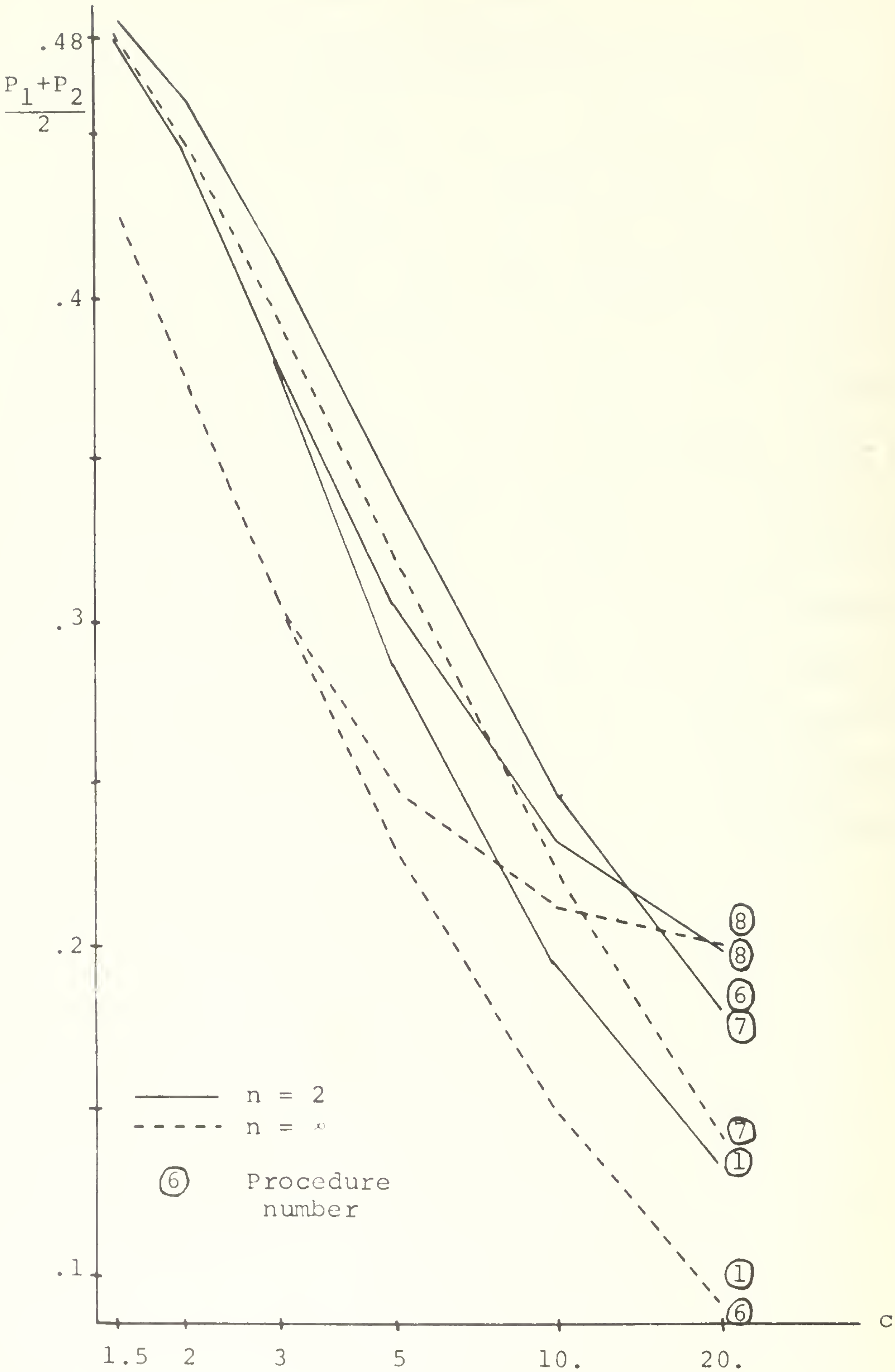

Figure 1

Expected Risk vs. c for various procedures; n = 2, ∞

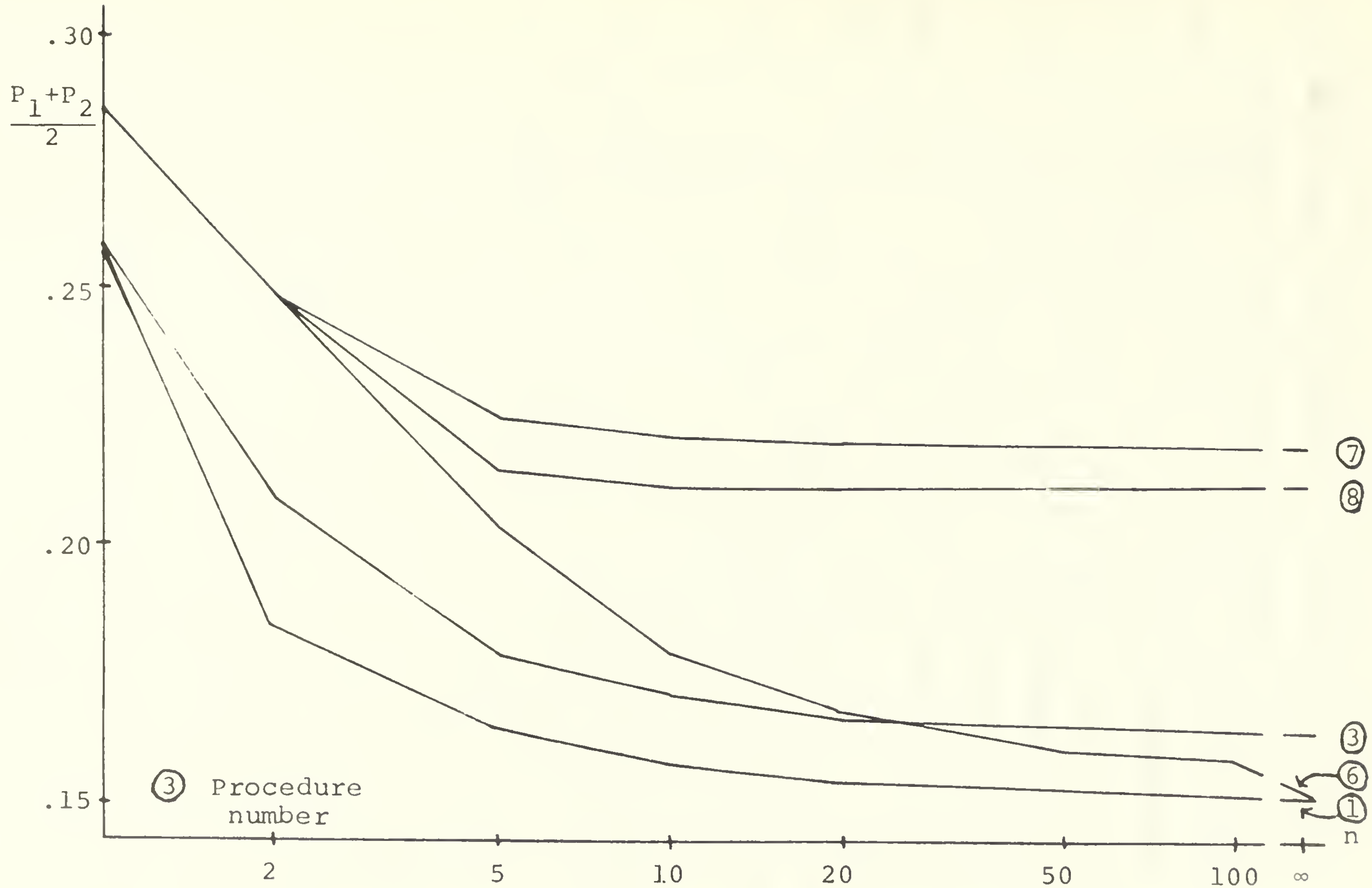

Figure 2

Expected Risk vs. n for Selected Procedures; c = 10

The negligible improvement in the performance of the likelihood ratio discriminant procedures for sample sizes in excess of 10 and the extremely slow approach to optimality of the Loftsgaarden and Quesenbury procedure are also interesting. An example of this for c = 10 is shown in Figure 2.

The considerable disparity of the values of P_1 and P_2 for many of the procedures considered in this thesis raises an interesting philosophical point which an investigator should settle for himself before selecting one of these methods for use. If, for example, one is willing to accept the possibility that a large percentage of the members of one population will be misclassified, although the average number of misclassifications is apt to be moderate, then the use of the linear discriminant function may be preferable to the use of the non-parametric procedures (unless c is very large). If, however, one is reassured by the fact that the rule of nearest neighbor makes errors no more than half the time (asymptotically) no matter what the situation, one may have a predilection for that procedure. The superiority in terms of expected risk of the linear discriminant function over the non-parametric procedures for small c is shown in Figure 1 where, for example for $n = 2$, $c = 5$ the linear discriminant has expected risk about .03 lower than the rule of nearest neighbor; for $n = \infty$, $c = 5$ the difference is almost .06. In fact, the performance of the linear discriminant where $c \leq 3$ is almost identical with that of the best procedure in this range, $L(1;\underset{\sim}{\lambda},\underset{\sim}{\mu})$. However, reference to Figure

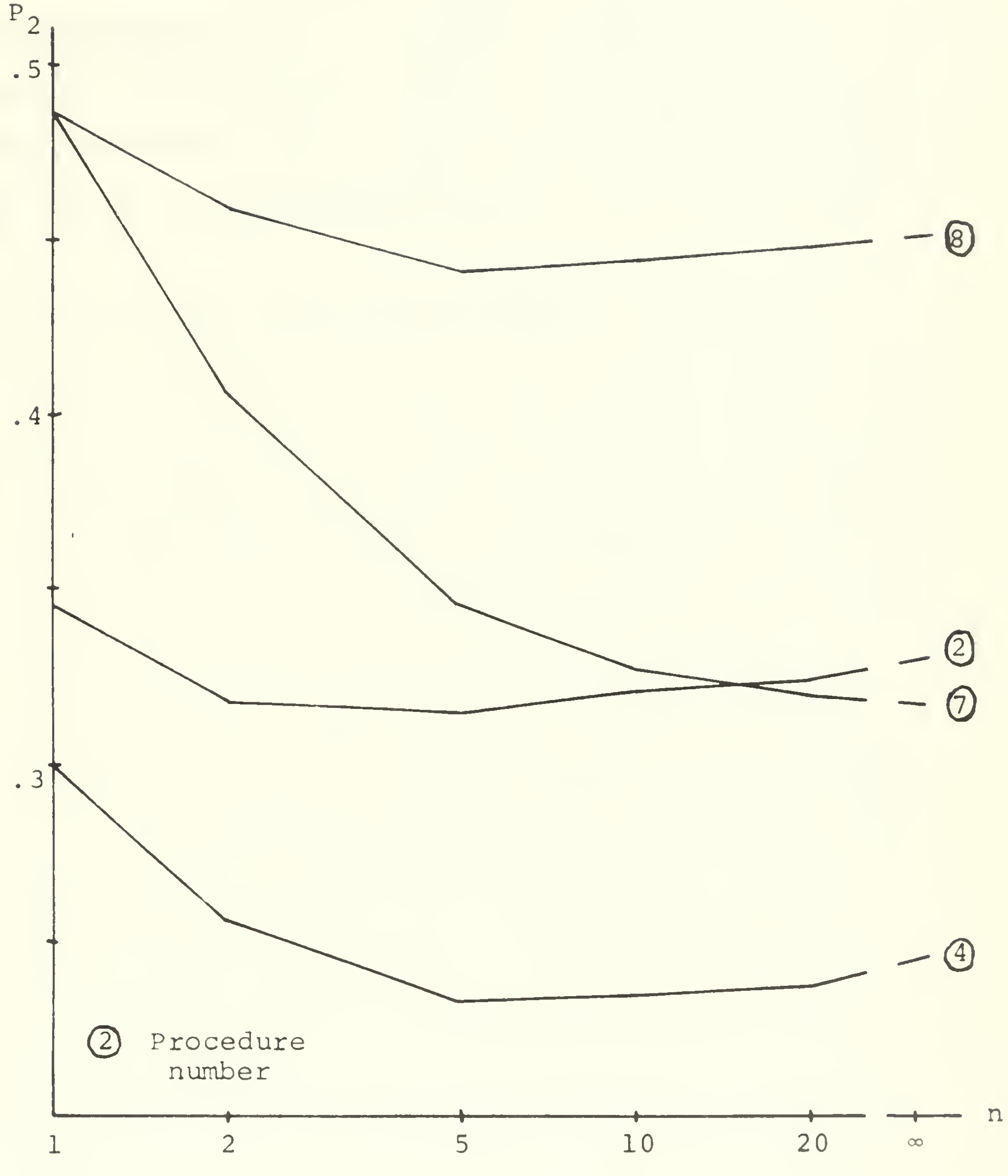

Figure 3

P_2 For Selected Procedures; c = 5

3 indicates that in the same cases, P_2 for the linear discriminant is much greater than that for the rule of nearest neighbor. Also apparent in Figure 3 is the non-monotonicity of P_2 for several procedures. Table 1 gives both P_1 and P_2 for all cases considered in this thesis so that expected risks for mixing probabilities other than $(\frac{1}{2},\frac{1}{2})$ may be easily calculated.

The following recommendations seem appropriate based on this study. If one can be reasonably certain that the populations are negative exponential, and there is no reason to suppose that the unknown observation is more likely to be from one of the populations than from the other, the minimax version of $L(t;\tilde{\lambda},\tilde{\mu})$ (Procedure 3) would be a good choice if $n \geq 5$. For smaller samples the same procedure with $t = 1$ (Procedure 1) seems better. If observations from one of the populations are appreciably more likely than those from the other, a procedure taking this fact into account by taking more observations from the more likely population and/or estimating the probability of occurrence of the populations (if these probabilities are not known) should be considered. A selection of the parameter t in the chosen procedure in order to minimize the expected risk with respect to the estimated (or known) population probabilities could then be made. Because the probability of classification error does not decrease appreciably as n increases from ten to infinity for the likelihood ratio procedures, it appears that the use of samples larger than ten in Procedures 1 - 4 is unwarranted

unless the cost of sampling is very small. If one cannot be certain that the populations are negative exponential, a choice between linear discriminant and a non-parametric procedure may be appropriate. The attitude of the experimenter toward the importance of P_1 and P_2 individually should influence his decision in this case.

BIBLIOGRAPHY

1. Cover, T.M., and P. E. Hart. Nearest Neighbor Pattern Classification, IEEE Transactions on Information Theory, IT-13: 21-27 (1967).

2. Fix, Evelyn, and J. L. Hodges, Jr. Nonparametric Discrimination. 1. Consistency Properties. Report prepared under contract between University of California and School of Aviation Medicine, Randolph Field,Texas. February 1951.

3. ________ . Nonparametric Discrimination: Small Sample Performance. Report prepared under contract between University of California and School of Aviation Medicine, Randolph Field,Texas. August 1952.

4. Hager, Charles F. A Comparison of Error Probabilities for Various Discrimination Procedures when the Populations are Gamma. Unpublished Master's thesis, Naval Postgraduate School, Monterey, California, 1964.

5. Hoel, P. G., and R. P. Peterson. A Solution to the Problem of Optimum Classification, Anals. Math. Stat.: XX. 433-8.

6. Loftsgaarden, D. O., and C. P. Quesenbury. Nonparametric Density Estimation and Classification. A report prepared under Grant No. NsG-562 by Montana State College, Bozeman, Montana for National Aeronautics and Space Administration (NASA TN D-26699).

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	20
2.	Library Naval Postgraduate School Monterey, California 93940	2
3.	Professor J. R. Borsting Department of Operations Analysis Naval Postgraduate School Monterey, California 93940	10
4.	Lieutenant J. L. Lockett III 676 Homer Avenue Palo Alto, California 94301	1
5.	Department of Operations Analysis Naval Postgraduate School Monterey, California 93940	1

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Naval Postgraduate School Monterey, California	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

A COMPARISON OF SOME PARAMETRIC AND NON-PARAMETRIC DISCRIMINATION PROCEDURES IN NEGATIVE EXPONENTIAL POPULATIONS

4. DESCRIPTIVE NOTES (Type of report and, inclusive dates)

Thesis

5. AUTHOR(S) (First name, middle initial, last name)

LOCKETT, Joseph Louis III, Lieutenant, USN

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
December 1968	45	6

8a. CONTRACT OR GRANT NO. b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Naval Postgraduate School Monterey, California

13. ABSTRACT

A brief discussion of the literature concerned with the two-population discrimination problem is presented and several procedures based on the likelihood ratio for discrimination between negative exponentially distributed populations are proposed. The small sample and asymptotic performance of these procedures is compared with that of non-parametric procedures and the classical linear discriminant function. Some guidelines for the use of the procedures discussed are presented.

UNCLASSIFIED
Security Classification

A-31408

KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
CLASSIFICATION						
NON-PARAMETRIC						
DISCRIMINATION						
EXPONENTIAL						

DD FORM 1 NOV 65 1473